

Universitäts-Spital Zürich
Department für Innere Medizin
Direktor: Prof. Dr. med. Andreas Schaffner
Abteilung für Naturheilkunde
Leiter: Prof. Dr. med. Reinhard Saller

Arbeit unter Leitung von Herrn PD Dr. med. Dieter Melchart

**Untersuchung zur Reliabilität der Zungendiagnose nach TCM
(Traditioneller Chinesischer Medizin)**

INAUGURAL-DISSERTATION
Zur Erlangung der Doktorwürde der Medizinischen Fakultät
der Universität Zürich

Vorgelegt von

Yanqing Li

aus München / Deutschland

Genehmigt auf Antrag von Prof. Dr. med Reinhard Saller
Zürich 2007

Inhaltsverzeichnis

1.	Einleitung und Fragestellung	3
2.	Hintergrundinformationen zur Traditionellen Chinesischen Medizin und zur Zungendiagnostik	5
2.1.	Grundlagen der Traditionellen Chinesischen Medizin	5
2.2.	Allgemeine Informationen zur Zungendiagnostik	6
2.3.	Anwendungsbeispiele	8
3.	Methodik	10
3.1.	Pilotstudien und Hauptstudie	10
3.2.	Design	10
3.3.	Setting	10
3.4.	Auswahl und Beurteilung der Zungenfotografien	10
3.4.1.	Auswahl der Zungenfotografien	10
3.4.2.	Beurteilung der Zungenmerkmale in Pilotstudie 1	11
3.4.3.	Beurteilung der Zungenmerkmale in Pilotstudie 2	13
3.4.4.	Beurteilung der Zungenmerkmale in der Hauptstudie	13
3.4.5.	Beurteiler	13
3.5.	Statistische Auswertung	16
4.	Ergebnisse	18
4.1.	Pilotstudie 1	18
4.2.	Pilotstudie 2	21
4.3.	Hauptstudie	23
5.	Diskussion	27
5.1.	Kurzzusammenfassung der Hauptergebnisse	27
5.2.	Diskussion der verwendeten Methoden	27
5.3.	Stand der Forschung zur Beschreibung der Zungenmorphologie im Rahmen der TCM	29
5.4.	Interpretation der Ergebnisse	31
5.5.	Schlussfolgerung und Empfehlungen für zukünftige Forschungsvorhaben	32
6.	Literaturverzeichnis	34
7.	Zusammenfassung	36
8.	Summary	38
	Danksagung	

1. Einleitung und Fragestellung

Komplementärmedizinische Verfahren erfreuen sich weltweit großer und zunehmender Popularität (Eisenberg et al., 1998; Härtel und Volger, 2004; Tindle et al., 2005). Die Traditionelle Chinesische Medizin (TCM) spielt hierbei eine besonders wichtige Rolle (WHO Global Atlas of Traditional, Complementary and Alternative Medicine, 2005). Während im Westen die Akupunktur das am weitesten verbreitete Therapieverfahren der TCM ist, spielt in China die Traditionelle Arzneimitteltherapie die vorherrschende Rolle. Wichtige weitere Therapien der TCM sind Tuina (Massage) und Qi Gong (Entspannung); darüber hinaus wird der Ernährung große Bedeutung beigemessen. Schätzungen zufolge erfolgt im heutigen China ein Drittel der medizinischen Versorgung der Bevölkerung unter Anwendung von TCM-Verfahren (Neeb, 2006). Bei der Befragung einer repräsentativen Stichprobe von 16.141 Personen der Schweizer Wohnbevölkerung mit einem Alter über 15 Jahren (Schweizer Gesundheitsbefragung) gaben 1,7% an, im Jahr 2002 mit Chinesischer Medizin behandelt worden zu sein (Crivelli et al, 2004). Inwieweit diese Zahl Akupunkturbehandlungen mit einschließt oder ausschließt, ist allerdings unklar.

Die TCM besitzt eine völlig eigenständige – aus naturwissenschaftlicher Sicht schwer nachvollziehbare – Konzeption (siehe Kapitel 2). Neben therapeutischen Verfahren hat sie auch eine eigene Systematik in der Diagnostik entwickelt. Die Beurteilung der Zunge spielt hierbei eine wichtige Rolle. Eine Evaluation, die sowohl westlichen wissenschaftlichen Ansprüchen, als auch der Theorie und Praxis der Traditionellen Chinesischen Medizin gerecht wird, ist nicht einfach zu bewerkstelligen. Aus wissenschaftlicher Sicht muss ein diagnostisches Verfahren reliabel und valide sein (Jensen und Abel, 1999). Als reliabel wäre die Zungendiagnose zu bewerten, wenn mehrere angemessen ausgebildete und erfahrene Ärzte bei der Beurteilung zur gleichen Bewertung kämen. Ein diagnostischer Test wird als valide bezeichnet, wenn er das Vorhandensein bzw. das Nicht-Vorhandensein einer Krankheit weitgehend sicher erkennt.

Insbesondere die Prüfung der Validität ist im Falle der Zungendiagnose (und vieler anderer komplementärer Diagnoseverfahren) problematisch. In der westlichen Medizin erfolgt die Prüfung der Validität durch einen Vergleich des jeweiligen Tests im Vergleich zum sogenannten Gold Standard, der besten und sichersten Möglichkeit eine Diagnose zu stellen (Fletcher und Fletcher, 2005). So würde man z.B. ein einfaches Screeninginstrument für das Vorhandensein einer Depression mit einer ausführlichen psychiatrischen Exploration vergleichen oder ein Belastungs-EKG zur Untersuchung einer koronaren Herzerkrankung mit einer Koronarangiographie. Im Falle der Zungendiagnose würde man untersuchen, inwieweit

eine alleinige Beurteilung der Zunge bereits die vollständige Syndromdiagnose vorhersagt bzw. inwieweit die Bewertung der Zunge zur Erstellung einer sicheren Syndromdiagnose beiträgt, im Vergleich zu einem diagnostischen Vorgehen ohne Zungendiagnose. Die Chinesische Syndromdiagnose, die eher eine funktionale Zustandsbeschreibung des Organismus unter Berücksichtigung konstitutioneller Aspekte darstellt, als eine „feste“ Diagnose im westlichen Sinne, eignet sich für eine solche Vorgehensweise nur eingeschränkt. Dementsprechend fehlt eine eindeutige diagnostische Klassifikation, welche die terminologischen Voraussetzungen für eine angemessene Validitätsprüfung schaffen würde.

Es erscheint einfacher, sich bei den ersten Schritten der Forschung auf die Prüfung von Aspekten der Reliabilität zu konzentrieren. Bei der Reliabilitätsprüfung sind zwei Ebenen denkbar. Zum einen kann man überprüfen, ob verschiedene Beurteiler die einzelnen morphologischen Merkmale von Zungen übereinstimmend *beschreiben*, z.B. ob mehrere Beurteiler die Zungenfarbe als auffällig rot bewerten. Eine solche Prüfung erscheint unproblematisch, ist jedoch nur von geringer klinischer Relevanz. Zum anderen kann man auf einer komplexeren Ebene untersuchen, ob die *Interpretation* der einzelnen Merkmale oder auch des Gesamtbildes der Zunge übereinstimmt. Diese Ebene besitzt deutlich größere klinische Relevanz. Für die hierfür notwendige terminologische Standardisierung sind jedoch umfangreiche Vorarbeiten und Vorprüfungen notwendig.

Literatursuchen in der Datenbank der US-amerikanischen National Library of Medicine PubMed (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?DB=pubmed>) deuten darauf hin, dass sich die Forschungsbemühungen im Zusammenhang mit der Zungendiagnostik in China auf die Digitalisierung und Automatisierung der Beurteilung von Zungenbildern konzentrieren (z.B. Chiu, 2000; Wei et al., 2002; Zhou et al., 2004). Untersuchungen zu Reliabilität oder Validität konnten nicht identifiziert werden. Über persönliche Kontakte konnten zwei, lediglich als Abstracts verfügbare, westliche Untersuchungen zu Aspekten der Reliabilität bekannt (Biltstein und Bonzak, 2006; Gareus et al., 2005) gefunden werden.

In der vorliegenden Dissertation zugrunde liegenden Studienserie sollten erste Schritte für eine wissenschaftlich angemessene Untersuchung der Reliabilität der Zungendiagnostik im Rahmen der TCM unternommen werden. Untersucht wurde die Frage, inwieweit unterschiedliche Beurteiler bei der morphologischen Beschreibung für die Zungendiagnostik wichtiger Merkmale übereinstimmen.

2. Hintergrundinformationen zur Traditionellen Chinesischen Medizin und zur Zungendiagnostik

2.1. Grundlagen der Traditionellen Chinesischen Medizin

Die TCM ist ein in sich geschlossenes, auf der chinesischen Naturphilosophie gegründetes Medizinsystem. Ihre Lehre wurzelt in der philosophischen Annahme, dass die Welt mit all ihren lebendigen Wesen darin einem ständigen Fluss und Wandel unterlegen ist, in dem letztlich alle Prozesse durch das Wirken universaler, funktioneller Prinzipien zustande kommen. Die TCM kennt keine eindeutige Trennung zwischen Körper und Geist, und der Organismus wird als Teil der Natur begriffen. Um einen Zustand der Gesundheit aufrecht zu erhalten, muss eine Harmonie der Körperfunktionen sowie des Körpers mit der Natur bestehen (Lao, 1999).

Die wichtigsten Theorien der TCM sind das Konzept des Yin und Yang und die Lehre der fünf Wandlungsphasen (Porkert, 1986). Das Konzept des Yin (wörtlich: die unbeleuchtete Seite eines Hügels) und Yang (die sonnige Seite eines Hügels) entspringt der taoistischen Naturphilosophie und besagt, dass sich alle Dinge in diese zwei Polaritäten teilen lassen. Alle Erscheinungen und Veränderungen des Lebens sowie alle Naturphänomene lassen sich demnach aus dem Wechselspiel von Yin und Yang ableiten (Blunck et al., 2006). Yin repräsentiert üblicher Weise die passiven, abnehmenden, inneren, kühlen und dunklen Aspekte eines Phänomens, Yang die aktiven, ansteigenden, äußeren, warmen und hellen. Sowohl die normalen wie auch die pathologischen Prozesse im Körper können aus Sicht der TCM nach Yin und Yang Charakteristiken differenziert werden (Lao, 1999). Störungen im Gleichgewicht des Yin und Yang führen zu Krankheiten. Symptome wie innere Unruhe, Fieber, Schlaflosigkeit oder Kopfschmerzen werden z.B. als Zeichen für einen Überfluss an Yang interpretiert, Symptome wie ein trockener Mund, trockene Haut oder Blutarmut als Zeichen einen Yin-Mangels.

Die Theorie der fünf Wandlungsphasen ordnet alle Naturphänomene den Kategorien Holz, Feuer, Erde, Metall und Wasser zu, und definiert deren Beziehungen zueinander (Blunck et al., 2006). Jeder der fünf Phasen werden spezifische Qualitäten zugeordnet. Der Yangphase Holz werden beispielsweise bei einem bereits aktiven Grundzustand weiter aufsteigende Tendenzen zugeordnet; als Jahreszeit wird der Frühling assoziiert. Dementsprechend wird Feuer dem Sommer, Erde dem Spätsommer, Metall dem Herbst und Wasser dem Winter zugeordnet. Entsprechende Zuordnungen finden sich für Organe, Emotionen, Geschmacksrichtungen, Farben etc. (Lao, 1999).

Darüber hinaus besitzt die TCM weitergehende konzeptionelle Vorstellungen, die zu einer eigenen Nosologie und entsprechenden Behandlungskonzepten führen. So gibt es eine Substanzlehre, eine Lehre zu so genannten Zang-Fu Funktionskreisen, Theorien zu einem Meridian- und Netzgefäßsystem, sowie definierte Vorstellungen zu pathogenen Faktoren (Blunck et al., 2006). Ein hier noch genannter, für die diagnostischen und therapeutischen Vorgehensweisen der TCM zentraler Begriff ist Qi. Das Qi ist eine von fünf „Grundsubstanzen“ (Kaptchuk, 1995) und wird häufig als Lebensenergie übersetzt.

Im Rahmen der theoretischen Konzeption der TCM werden pathologische Zustände als Syndrome klassifiziert (Lu et al. 2004). Patienten, die aus Sicht der westlichen Medizin eine definierte Diagnose haben, können in der TCM an unterschiedlichen Syndromen leiden, gleichzeitig können zwei Patienten mit unterschiedlichen westlichen Diagnosen das gleiche Syndrom haben.

Aus der Sicht skeptischer Wissenschaftler sind die Theorien der TCM naturwissenschaftlich nicht haltbar (z.B. Renckens, 2002). Darüber hinaus wird aus medizinhistorischer Sicht kritisch darauf hingewiesen, dass die TCM keine einheitliches, historisch kontinuierlich gewachsenes System ist, sondern äußerst heterogen sei und insbesondere nach der Kulturrevolution in China politisch stark instrumentalisiert wurde (Unschuld, 2003)

2.2. Allgemeine Informationen zur Zungendiagnostik

Die grundlegenden diagnostischen Methoden der TCM sind die Anamnese, die Inspektion, Auskultation und Olfaktion sowie die Palpation (Maret et al., 2006). Die chinesische Diagnose soll nicht nur Aufschluss über Art und Lokalisation der Störung geben, sondern auch eine Aussage über die Konstitution und aktuelle Kondition des Patienten beinhalten, d.h. sie beurteilt auch den vegetativen Grundtonus des Patienten und seine aktuelle Reaktionslage.

Die Zungendiagnostik ist eine spezielle Technik im Rahmen der Inspektion. Eine vollständige TCM-Diagnose ist ohne Zungendiagnostik nicht möglich, gleichzeitig wird betont, dass die Zunge nicht notwendiger Weise das gesamte Syndrom, sondern unter Umständen nur einzelne Symptome und damit einen Teil des Syndroms, widerspiegelt (Maret et al., 2006). Einzelne TCM-Experten vertreten allerdings die Ansicht, dass ca. 90 Prozent der für die

Syndromdiagnose relevanten Krankheitszeichen und Krankheitsveränderungen durch die Zungendiagnose erkenn- und beurteilbar seien (Yuan et al., 2002). Auch in deutscher Sprache liegen umfangreiche Lehrbücher und Atlanten speziell zur Zungendiagnose vor (z.B. Maciocia, 2000; Kirschbaum, 2002).

Die TCM geht davon aus, dass die Zunge über Meridiane und Netzgefäße mit den bereits oben genannten Zung-Fu-Organen verbunden ist und Prozesse im Organismus sich über diese Verbindungen auf die Zunge auswirken. So wird z.B. postuliert, dass der Blasen-Meridian sich auf der Zunge verzweigt, während der Nieren-Meridian in der Zungenwurzel endet (Maret et al., 2006). Die Inspektion der Zunge wird nicht nur zur Diagnosestellung als hilfreich erachtet sondern auch als wichtiges Kriterium bei der Verlaufbeurteilung herangezogen.

Die Inspektion der Zunge soll möglichst bei hellem Tageslicht erfolgen. Wenn dies nicht möglich ist, sollte eine künstliche Beleuchtung mit Tageslichtlampen verwendet werden (Yuan, 1996). Zunächst wird der Patient gebeten, die Zunge langsam ohne Verkrampfung soweit wie möglich aus dem Mund herauszustrecken. Die Zungenspitze soll dabei etwas nach unten geneigt sein. Anschließend wird der Zungenbelag überprüft, seine Farbe und seine Stärke, bzw. seine Dicke. Anschließend erfolgt eine sorgfältige Examinierung von der Zungenspitze über die Zungenmitte und die Zungenränder bis hin zur Zungenwurzel. Nach der Inspektion des Zungenbelages wird versucht, die Qualität der Zunge selbst zu beurteilen. Dabei geht es vor allem um die Begutachtung der Zungenfarbe, wie z.B. hell, blass, rot, tiefrot, purpurn oder blau. Rote Punkte, blaue Flecken oder kleine Geschwüre werden als deutliche Hinweise sorgfältig festgehalten. Um bei einer dicht und dick belegten Zunge die richtige Zungenfarbe zweifelsfrei zu erkennen, wird der Zungenbelag z.B. mit Hilfe eines Wattestäbchens entfernt. Weiterhin werden Form, Größe und Beweglichkeit der Zunge beurteilt.

Bei der Inspektion der Zunge muss auf einige wichtige Faktoren geachtet werden, welche unter Umständen die Diagnose verfälschen könnten (Maret et al., 2006). So sollte die Inspektion entweder in nüchternem Zustand oder ca. 2 Stunden nach der Nahrungsaufnahme vorgenommen werden, da sich der Zungenbelag durch das Kauen von dick zu dünn verändern kann. Je mehr getrunken wird, umso feuchter ist die Zunge. Wird weniger getrunken, dann zeigt sich ein tendenziell trockener Zungenbelag. Nimmt der Patient heiße und scharfe Nahrung zu sich, ist die Zungenfarbe tiefer rot. Einzelne Nahrungsmittel können die Farbe des Zungenbelages ebenso verändern (z.B. Rotfärbung durch rote Beete, Gelbfärbung durch Karotten oder schmierig-weiße Beläge durch Verzehr von Bananen) wie

Arzneimittel (z.B. Vitamin B, einzelne Kräutertees oder Crataegus-Präparate) oder Getränke (z.B. Milch, Coca Cola, Limonaden oder Orangensaft). Nehmen die Patienten über längere Zeit Tee, Kaffee oder Nikotin zu sich, wird der Zungenbelag braun und trocken. Daher sollten bei auffälligen Befunden derartige Ursachen ausgeschlossen werden (Yuan, 1996).

2.3. Anwendungsbeispiele

Um Vorgehensweisen, Beschreibung und Interpretation bei der Zungendiagnostik zu veranschaulichen werden im folgenden 4 Beispiele präsentiert (siehe Abbildung 1). Die Zunge im linken oberen Feld entspricht aus Sicht der TCM einer normalen, gesunden Zunge. Sie hat einen dünnen, weißen Belag, ist leicht feucht und hat eine leicht rote Farbe. Die Zungengröße ist unauffällig.

Bei der Zunge im rechten oberen Feld ist fast kein Belag zu erkennen, die Zunge ist ebenfalls leicht feucht, hat jedoch eine leicht blasse Farbe. Sie erscheint geschwollen und zeigt multiple, in unterschiedliche Richtungen verlaufende Risse und Furchen. Tiefe Risse und Furchen gehen häufig mit einer schmerzenden Zunge einher. Differentialdiagnostisch sind die Risse und Furchen aus Sicht der TCM als Zeichen eines „Blut-Mangels“ oder „Yin-Verlustes“ zu interpretieren, der häufig konstitutionell bedingt ist. Bei einer geschwollenen, hellweißen Zunge wäre an eine „Yang-Leere von Milz und Magen“ zu denken. Zu beachten ist allerdings, dass etwa 0,5% aller gesunden Menschen seit der Geburt Risse und Furchen in der Zunge haben, bei denen diese Zeichen nicht als Hinweis auf ein pathologisches Geschehen bewertet würden. Die Zungenzeichen weisen in ihrer Gesamtheit auf eine Syndromdiagnose „Blut-Schwäche durch Nässe-Kälte“ hin.

Die Zunge im linken unteren Feld ist von einem dicken, leicht gelben, trockenen Belag im gesamten Bereich bedeckt. Die Farbe der Zunge erscheint blass, die Zunge insgesamt geschwollen. An den Seiten sind deutliche Zahneindrücke zu erkennen, im mittleren Bereich finden sich Risse. Differentialdiagnostisch ist eine blasse, geschwollene Zunge mit Zahneindrücken ein Zeichen für einen „Milz-Yang-Mangel“ oder eine „Qi (Yang) Schwäche“. Die leicht gelbe Farbe des Zungenbelags würde darauf hinweisen, dass der „Yang-Mangel“ durch „Hitze“ verursacht ist. Die Risse im mittleren Bereich deuten dagegen auf einen „Magen-Yin-Mangel“ hin. Insgesamt spricht die Zunge dafür, dass der Patient an einem „Magen-Yin-Mangel“ wegen „Milz-Yang-Mangels durch Hitze“ leidet.



Abbildung 1

Zungen von vier Patienten mit unterschiedlichen Zeichen (Beschreibung im Text)

Bei der Zunge rechts unten fällt im Zungenspitzenbereich ein dicker, leicht gelber Belag auf; in der Zungenmitte (Zuordnung nach TCM zu Magen und Milz) und im Wurzelbereich (Zuordnung Niere und Dickdarm) ist die Farbe dunkel-gelb. Der Belag wirkt eher trocken. Die Zungenfarbe ist leicht bläulich, die Zunge erscheint insgesamt relativ groß. Differentialdiagnostisch spricht eine dunkel-gelbe Farbe im mittleren Bereich und im Wurzelbereich für eine „Hitze-Krankheit im Magen-Dickdarm-Bereich“, ein „Inneres Krankheitssyndrom“ sowie ein „Hitze-Syndrom.“ Je dunkler die gelbe Farbe ist, umso stärker ist aus Sicht der TCM die Hitzestörung. Ein rauchfarbenes Gelb, das ins bräunliche übergeht, deutet auf eine „gestaute äußere Hitzestörung“ hin. Eine leicht bläuliche Zungenfarbe ist eine Hinweis für eine leichte „Blut-Hitze“ Störung. Die Diagnose auf Basis der Zungenzeichen wäre eine „Hitze-Krankheit“ und „Fülle-Hitze im Magen-Dickdarm.“

3. Methodik

3.1. Pilotstudien und Hauptstudie

Im Rahmen der Dissertation wurden zwei Pilotstudien und eine Hauptstudie durchgeführt. Ziel der ersten Pilotstudie (Pilotstudie 1) war die Erprobung der Methodik, d.h. die praktische Erprobung des entwickelten Beurteilungsbogens, das Ermitteln von Problemen bei der Beurteilung sowie der Erhalt erster orientierender Ergebnisse. Ziel der zweiten Pilotstudie (Pilotstudie 2) war die Erprobung eines aufgrund der Erfahrungen von Pilotstudie 1 überarbeiteten Beurteilungsf formulars, das nach Vorabsprache über die Bewertung von Grenzbefunden eingesetzt wurde. Nach der Vortestung der Methodik in den beiden Pilotstudien sollte in der Hauptstudie die Reliabilität der Beschreibung der morphologischen Zungencharakteristika bei einer größeren Zahl von Zungenfotografien getestet werden.

3.2. Design

Beide Pilotstudien und die Hauptstudie wurden durchgeführt als standardisierte, deskriptive Beurteilungen von Zungenmerkmalen durch drei (erste Pilotstudie und Hauptstudie) bzw. zwei (zweite Pilotstudie) unabhängige, verblindete Beurteiler.

3.3. Setting

Die Durchführung der Studien erfolgte in der Klinik für Traditionelle Chinesische Medizin in Kötzing. In dieser Klinik mit 76 Betten werden pro Jahr ca. 1000 Patienten mit meist chronischen Erkrankungen stationär (mittlere Liegedauer 26 Tage) von 10 chinesischen Ärzten und 6 deutschen Ärzten primär mit den Methoden der Traditionellen Chinesischen Medizin (Akupunktur, chinesische Arzneimitteltherapie, Tuina und Qi Gong) behandelt.

3.4. Auswahl und Beurteilung der Zungefotografien

3.4.1. Auswahl der Zungenfotografien

Zwischen 2003 und 2005 wurde routinemäßig von allen stationären Patienten der Klinik am Morgen nach der Aufnahme vor dem Frühstück im gleichen Raum unter gleichen Lichtbe-

dingungen von einer deutschen Laborkraft mit einer Digitalkamera (*Fuji Film Smart Media, Digital Camera, Fine Pix 2600 Zoom 3V*) eine Fotografie der Zunge erstellt. Die Patienten saßen dabei immer auf dem gleichen Stuhl, um den Winkel, in dem fotografiert wurde, möglichst konstant zu halten. Die Fotografien wurden digital und kodiert in einer elektronischen Patientenakte abgelegt (Melchart, 2001). Für die Untersuchung wurden von einer am weiteren Untersuchungsverlauf nicht beteiligten Verwaltungskraft aus dem Pool von Foto-Dateien zufällig Fotografien gezogen. Für die beiden Pilotstudien wurden jeweils 15 Fotografien gezogen, für die Hauptstudie 101 Fotografien. Die Fotografien wurden für die Beurteiler in einzelne Word-Dateien kopiert; die Dateien waren fortlaufend nummeriert. Für die Beurteiler war eine Zuordnung der Zungenfotografien zu Patienten nicht möglich. Für die zweite Pilotstudie und die Hauptstudie wurde von der Verwaltungskraft zusätzlich Alter, Geschlecht und Hauptdiagnose der jeweiligen Patientendokumentation ins Medical Monitoring eingegeben. Diese Information wurde separat abgelegt und stand den beurteilenden Ärzten bei der Zungenbewertung nicht zur Verfügung.

3.4.2. Beurteilung der Zungenmerkmale in Pilotstudie 1

Routinegemäß erfolgte in der Klinik eine teilstandardisierte Beschreibung der Zungenmerkmale durch einen chinesischen Arzt in einer elektronischen Patientenakte. Da die Dokumentation in dieser Patientenakte, wie sich in einem Probelauf gezeigt hatte, jedoch für die Studie zu viele Freiräume ließ (z.B. komplexe Kombination von Merkmalen sowie bei Nichtausfüllen eines Items Unklarheit, ob das Merkmal nicht vorlag oder lediglich klinisch nicht relevant war), wurde in Absprache mit den leitenden chinesischen und deutschen Ärzten der Klinik ein Beurteilungsformular (siehe Abbildung 2) erarbeitet. In Item 1 wurde die Zungenfarbe abgefragt. Aus Gründen der besseren Auswertbarkeit (siehe 3.2.4.) wurden vier Unterfragen (Zungefarbe rot, blass, bläulich oder rötlich-purpur) mit den Antwortoptionen „ja“ und „nein“ gestellt. In Item 2 wurde das Vorhandensein von Flecken beurteilt. Zungenform und andere Eigenschaften wurden in Item 3 in drei Unterfragen mit den Antwortoptionen „ja“ und „nein“ (Zunge gedunsen, Zahneindrücke vorhanden, Risse vorhanden) bewertet. Die vier Unterfragen von Item 4 betrafen die Art des Zungenbelags (dicker Belag, klebriger Belag, sich auflösender Belag, dünner Belag). Bei Item 5 zur Farbe des Zungenbelags wurden drei Antwortmöglichkeiten vorgegeben (gelb, weiß, andere Farbe bzw. nicht anwendbar). Bei der Auswahl der Items und der Antwortmöglichkeiten war neben einer einfachen Auswertbarkeit die Relevanz der Merkmale aus Sicht bzw. auf Basis der persönlichen Erfahrung der chinesischen und deutschen Ärzte das wesentliche Kriterium.

Keiner der Ärzte hatte Vorerfahrung mit der formalen Untersuchung der Zungendiagnostik im Rahmen von Studien.

Bild-Nr: _____	Beurteiler: _____
1. Zungenfarbe	
a) Farbe rot? o ja o nein	b) Farbe blass? o ja o nein
c) Farbe bläulich? o ja o nein	d) Farbe rötlich-purpur o ja o nein
2. Flecken	
o ja o nein	
3. Zungenform	
a) Zunge gedunsen? o ja o nein	b) Zahneindrücke vorhanden? o ja o nein
c) Risse vorhanden? o ja o nein	
4. Zungenbelag - Qualität	
a) Dicker Belag? o ja o nein	b) Klebriger Belag? o ja o nein
c) Sich auflösender Belag? o ja o nein	d) Dünner Belag o ja o nein
5. Zungenbelag - Farbe	
o gelb o weiß o andere Farbe/nicht anwendbar	

Abbildung 2

Beurteilungs-Formular für die Pilotstudie 1

3.4.3. Beurteilung der Zungenmerkmale in Pilotstudie 2

Aufgrund der Ergebnisse der ersten Pilotstudie wurde das Vorgehen bei der Beurteilung der Zungenmerkmale für die Pilotstudie 2 modifiziert. Um die starke Beeinflussung des Farbeindrucks auf dem Computerbildschirm durch den Blickwinkel des Beurteilers (die Farbe erscheint auf dem Bildschirm umso dunkler, je höher die Augen des Beobachters liegen, d.h. einem größeren oder höher sitzendem Beobachter erscheint die Zunge dunkler als einem kleiner bzw. niedrig sitzendem) zu vermindern, wurde ein Gestell gebaut, auf das alle Beurteiler zur Bewertung der Bilder ihr Kinn auflegten. Damit wurde erreicht, dass der Blickwinkel für alle Beurteiler praktisch identisch war. Als weiteres Problem bei der Bewertung der Farbe erwies sich die Tatsache, dass diese in unterschiedlichen Regionen der Zunge variieren kann. Für die Beurteilung aus Sicht der TCM ist es von Bedeutung, ob ein auffälliger Farbbefund z.B. die Zungenspitze oder deren Ränder betrifft. Das Formular wurde daher entsprechend überarbeitet (siehe Abbildung 3). Ein drittes Hauptproblem in der Pilotstudie 1 war, dass unklare Befunde von den Beurteilern unterschiedlich bewertet wurden. Während ein Beurteiler z.B. nur pathologisch relevante Zahneindrücke als „Zahneindrücke vorhanden“ bewertete, zählte ein anderer Beobachter auch weniger deutliche, aber sichtbare Zahneindrücke. Vor der Bewertung der 15 Fotografien der Pilotstudie 1 erfolgte daher eine Abstimmung zwischen den Beurteilern auf der Basis der gemeinsamen Durchsicht der 15 Bilder der Pilotstudie 1.

3.4.4. Beurteilung der Zungenmerkmale in der Hauptstudie

Die Beurteilung der Fotografien in der Hauptstudie erfolgte wie in Pilotstudie 2. Es wurde lediglich eine Vereinfachung des Formulars bei Item 1 (siehe Abbildung 4) Zungenfarbe verwirklicht, da einzelne in der Pilotstudie 2 abgefragte Merkmale extrem selten auftreten und daher praktisch von geringer Relevanz erschienen.

3.4.5. Beurteiler

Die Zungenfotos wurden von insgesamt vier Beurteilern streng unabhängig voneinander (Blindbedingungen) mit Hilfe der Formulare bewertet. Die Bewertung erfolgte ausschließlich auf Basis der Zungenfotos am Computer-Bildschirm ohne jegliche weitere Informationen. Beurteiler 1 (JD) war der Chefarzt der chinesischen Delegation. Er verfügt über ein abgeschlossenes fünfjähriges Studium der TCM an der Universität Peking und 18 Jahre

klinische Erfahrung. Beurteiler 2 (JZ) hat ebenfalls ein fünfjähriges TCM-Studium absolviert und 21 Jahre klinische Erfahrung. Beurteilerin 3 (YL) war die Doktorandin (5 Jahre schulmedizinische Ausbildung und TCM-Aufbauausbildung, Erfahrung 15 Jahre). Beurteiler 4, war

1. Zungenfarbe		
Zungenspitze	Zungenränder	Zungenkörper
a) Farbe rot?	d) Farbe rot?	g) Farbe rot?
<input type="radio"/> ja <input type="radio"/> nein	<input type="radio"/> ja <input type="radio"/> nein	<input type="radio"/> ja <input type="radio"/> nein
b) Farbe blass?	e) Farbe blass?	h) Farbe blass?
<input type="radio"/> ja <input type="radio"/> nein	<input type="radio"/> ja <input type="radio"/> nein	<input type="radio"/> ja <input type="radio"/> nein
c) Farbe bläulich?	f) Farbe bläulich	i) Farbe bläulich
<input type="radio"/> ja <input type="radio"/> nein	<input type="radio"/> ja <input type="radio"/> nein	<input type="radio"/> ja <input type="radio"/> nein
2. Flecken		
<input type="radio"/> ja <input type="radio"/> nein		
3. Zungenform		
a) Zunge gedunsen?	b) Zahneindrücke vorhanden?	
<input type="radio"/> ja <input type="radio"/> nein	<input type="radio"/> ja <input type="radio"/> nein	
c) Risse vorhanden?		
<input type="radio"/> ja <input type="radio"/> nein		
4. Zungenbelag - Qualität		
a) Dicker Belag?	b) Klebriger Belag?	
<input type="radio"/> ja <input type="radio"/> nein	<input type="radio"/> ja <input type="radio"/> nein	
c) Sich auflösender Belag?	d) Dünner Belag	
<input type="radio"/> ja <input type="radio"/> nein	<input type="radio"/> ja <input type="radio"/> nein	
5. Zungenbelag - Farbe		
<input type="radio"/> gelb <input type="radio"/> weiß	<input type="radio"/> andere Farbe/nicht anwendbar	

Abbildung 3

Beurteilungs-Formular für die Pilotstudie 2

1. Zungenfarbe

Zungenspitze

a) Farbe rot?

☐ ja ☐ nein

Zungenränder

b) Farbe rot?

☐ ja ☐ nein

Zungenkörper

c) Farbe rot?

☐ ja ☐ nein

d) Farbe blass?

☐ ja ☐ nein

e) Farbe bläulich

☐ ja ☐ nein

2. Flecken

☐ ja ☐ nein

3. Zungenform

a) Zunge gedunsen?

☐ ja ☐ nein

b) Zahneindrücke vorhanden?

☐ ja ☐ nein

c) Risse vorhanden?

☐ ja ☐ nein

4. Zungenbelag - Qualität

a) Dicker Belag?

☐ ja ☐ nein

b) Klebriger Belag?

☐ ja ☐ nein

c) Sich auflösender Belag?

☐ ja ☐ nein

d) Dünner Belag

☐ ja ☐ nein

5. Zungenbelag - Farbe

☐ gelb ☐ weiß

☐ andere Farbe/nicht anwendbar

Abbildung 4

Beurteilungs-Formular für die Hauptstudie

der deutsche Chefarzt der TCM-Klinik Kötzing, der über zehn Jahre TCM-Erfahrung verfügt. In der Pilotstudie 1 erfolgte die Bewertung durch die drei Beurteiler 1, 2 und 3, in der Pilotstudie 2 durch die Beurteiler 1 und 2 und in der Hauptstudie durch die Beurteiler 1, 2 und 4.

3.5. Statistische Auswertung

Die erhobenen Daten wurden von der Doktorandin in eine Excel-Datei eingegeben (Einfacheingabe mit Plausibilitäts- und Stichprobenchecks), in SPSS eingelesen und statistisch ausgewertet. Für jedes Beurteilerpaar wurden die prozentuale Übereinstimmung ($((\text{Zahl der gleichen Bewertungen} / \text{Gesamtzahl der Bewertungen}) \times 100)$) sowie als Maß für die überzufällige Übereinstimmung der Kappa-Index nach Cohen und dessen statistische Signifikanz berechnet.

Die prozentuale Übereinstimmung der Bewertung eines diagnostischen Merkmals von zwei Beurteilern als vorhanden oder nicht vorhanden ist die einfachste Maßzahl zur Beschreibung der Übereinstimmung. Diese berücksichtigt jedoch nicht, dass ein Teil dieser Übereinstimmung zufällig ist bzw. sein kann. Werden z.B. zwei Münzen geworfen, erwartet man in 50% der Würfe, dass beide übereinstimmend Kopf oder Zahl zeigen. Daher wird der Kappa-Index nach Cohen (Fleiss, 1981) als Maß zur Quantifizierung der Übereinstimmung jenseits des Zufalls herangezogen. Dieses Maß berücksichtigt außerdem auch, ob eine Merkmal (z.B. eine auffällig rote Zunge) sehr häufig oder sehr selten auftritt. Die Berechnung des Kappa-Indexes erfolgt in drei Schritten. Im ersten Schritt wird die tatsächlich beobachtete Übereinstimmung ermittelt. In der folgenden Beispiel-Vier-Feldertafel beträgt diese 0,9 oder 90% (in 5 von 100 Fällen sagen beide Beurteiler die Zunge sei „rot,“ in weiteren 85 sagen beide, sie sei „nicht rot“).

Beispiel mit (hypothetischen) beobachteten Häufigkeiten

	Beurteiler 2		Summe
	Rot	Nicht rot	
Beurteiler 1			
Rot	5	5	10
Nicht rot	5	85	90
Summe	10	90	100

Im zweiten Schritt werden für die vier Kombinationsmöglichkeiten die erwarteten Häufigkeiten berechnet, indem die jeweilige Zeilensumme mit der Spaltensumme multipliziert und dann durch die Gesamtzahl an beurteilten Fällen geteilt wird. Für die Beurteilung „rot“ durch beide Beurteiler ist die Zeilensumme 10, die Spaltensumme ebenso 10; die bei nur zufälliger Übereinstimmung erwartete Häufigkeit der Bewertung „rot“ durch beide Beurteiler wäre demnach 1 ($10 \times 10 / 100$). Entsprechend ergibt sich für die beiden Nichtübereinstimmungsmöglichkeiten jeweils die erwartete Häufigkeit 9 ($10 \times 90 / 100$) und für das Urteil „nicht

rot“ durch beide Beurteiler die erwartete Häufigkeit 81 ($90 \times 90 / 100$). Die erwartete Gesamtübereinstimmung beträgt somit 0,82 bzw. 82%.

Beispiel mit den erwarteten Häufigkeiten in Klammern

Beurteiler 1	Beurteiler 2		Summe
	Rot	Nicht rot	
Rot	5 (1)	5 (9)	10
Nicht rot	5 (9)	85 (81)	90
Summe	10	90	100

Im dritten Schritt wird der eigentliche Kappa-Index berechnet, indem die erwartete von der beobachteten Übereinstimmung abgezogen ($0,9 - 0,82 = 0,08$) und dann durch $1 - \text{erwartete Beobachtung}$ ($1 - 0,82 = 0,18$) geteilt wird. Das errechnete Kappa für das Beispiel ist somit $0,08 / 0,18 = 0,44$. Kappawerte $< 0,40$ werden als geringe überzufällige Übereinstimmung klassifiziert, Werte zwischen 0,40 und 0,59 als moderate Übereinstimmung, Werte zwischen 0,60 und 0,74 als gute und Werte von mindestens 0,75 als sehr gute Übereinstimmung (Fleiss, 1981).

Der Kappa-Index kann nicht berechnet werden, wenn a) drei Felder der Vier-Felder-Tafel leer bleiben (z.B. in allen Fällen in denen beide Beurteiler alle Zungen als „nicht rot“ bewerten) oder b) wenn zwei Felder der Vier-Felder-Tafel leer bleiben und keine völlige Übereinstimmung herrscht (z.B. wenn beide Beurteiler in 99 Fällen die Zungen als „nicht rot“ bewerten, in einem Fall jedoch nicht übereinstimmen). Dagegen kann Kappa berechnet werden, wenn zwei Felder frei bleiben, jedoch 100% Übereinstimmung besteht (beide Beurteiler bewerten 99 Zungen als „nicht rot“ und eine Zunge als „rot“).

Die Standardeinstellungen bei der Berechnung von Kappa in SPSS beinhalten auch die Bestimmung eines p-Wertes. Für die Pilotserie wurde aus Praktikabilitätsgründen eine Zahl von 15 Fotografien ausgewählt. Eine formale Fallzahlschätzung erfolgte nicht. Aufgrund der kleinen Fallzahl und der großen Zahl von Testen werden die p-Werte in den Pilotserien 1 und 2 zwar in den Tabellen berichtet, jedoch nicht weitergehend diskutiert und interpretiert.

4. Ergebnisse

4.1. Pilotstudie 1

Die Übereinstimmung variierte zwischen den einzelnen Bewertungskriterien und zum Teil zwischen den einzelnen Beurteilern für ein einzelnes Kriterium deutlich (siehe Tabellen 1-3). Beim Vergleich der Beurteiler 1 und 2 schwankten die Kappa-Werte zwischen -0,15 und 0,76, die prozentuale Übereinstimmung zwischen 27% und 93%. Lediglich beim Item „Auflösender Belag“ ergab sich eine sehr gute Übereinstimmung über den Zufall hinaus (0,76), bei keinem Item eine gute Übereinstimmung (Kappa zwischen 0,60 und 0,74), bei den

Tabelle 1

Ergebnisse der ersten Pilotstudie (n = 15): Beurteiler 1 vs. Beurteiler 2

Item		Beurteiler 1 vs. 2		%	Kappa	Signifikanz p-Wert
		Nein	Ja			
Zungenfarbe rot	nein	1	1	47%	-0,02	0,92
	ja	7	6			
Zunge Blass	nein	11	2	73%	-0,15	0,55
	ja	2	0			
Zungenfarbe bläulich	nein	3	5	53%	0,08	0,71
	ja	2	5			
Zungenfarbe rötlich- purpur	nein	14	1	93%	n.b.	n.b.
	ja	0	0			
Flecken	nein	4	11	27%	n.b.	n.b.
	ja	0	0			
Gedunsen	nein	7	0	73%	0,48	0,03
	ja	4	4			
Zahneindrücke	nein	2	0	40%	0,11	0,36
	ja	9	4			
Risse	nein	9	0	67%	0,19	0,20
	ja	5	1			
Dicker Belag	nein	11	0	73%	n.b.	n.b.
	ja	4	0			
Klebriger Belag	nein	5	1	73%	0,47	0,06
	ja	3	6			
Auflösender Belag	nein	12	0	93%	0,76	0,002
	ja	1	2			
Dünner Belag	nein	0	4	73%	n.b.	n.b.
	ja	0	11			

Übereinstimmungen fettgedruckt; n.b. = nicht berechenbar (bei zwei oder mehr leeren Feldern in der Vierfeldertafel)

Items „Zunge gedunsen“ und „Klebriger Belag“ eine moderate Übereinstimmung (0,48 bzw. 0,47). Bei den übrigen Merkmalen lag der Kappa-Index unter 0,40. Für 4 Items konnte kein Kappa berechnet werden. Beim Vergleich der Beurteiler 1 und 3 war die prozentuale Übereinstimmung (40% - 100%; siehe Tabelle 2) meist höher als beim Vergleich der Beurteiler 1 und 2 (27-93%). Die Kappawerte schwankten wiederum stark (zwischen -0,13 und 0,73%). Bei keinem Item ergab sich eine sehr gute Übereinstimmung über den Zufall hinaus (Kappa-Wert größer als 0,74), bei „Zunge gedunsen“ und bei „Klebriger Belag“ eine gute Übereinstimmung (0,73 und 0,71), bei drei weiteren Merkmalen eine moderate Übereinstimmung („Klebriger Belag“ 0,47, „Auflösender Belag“ 0,44 und „Risse“ 0,41). Für drei Merkmale konnte kein Kappa berechnet werden.

Tabelle 2

Ergebnisse der ersten Pilotstudie (n = 15): Beurteiler 1 vs. Beurteiler 3

Item		Beurteiler 1 vs. 3		%	Kappa	Signifikanz p-Wert
		Nein	Ja			
Zungenfarbe rot	nein	2	0	40%	0,11	0,36
	ja	9	4			
Zunge Blass	nein	7	6	60%	0,24	0,15
	ja	0	2			
Zungenfarbe bläulich	nein	7	1	47%	-0,13	0,33
	ja	7	0			
Zungenfarbe rötlich- purpur	nein	15	0	100%	n.b.	n.b.
	ja	0	0			
Flecken	nein	14	1	93%	n.b.	n.b.
	ja	0	0			
Gedunsen	nein	5	2	87%	0,73	0,003
	ja	0	8			
Zahneindrücke	nein	2	0	67%	0,30	0,10
	ja	5	8			
Risse	nein	8	1	73%	0,41	0,09
	ja	3	3			
Dicker Belag	nein	9	2	87%	0,71	0,004
	ja	0	4			
Klebriger Belag	nein	5	1	73%	0,47	0,06
	ja	3	6			
Auflösender Belag	nein	12	0	87%	0,44	0,04
	ja	2	1			
Dünner Belag	nein	0	0	87%	n.b.	n.b.
	ja	2	13			

Übereinstimmungen fettgedruckt; n.b. = nicht berechenbar (bei zwei oder mehr leeren Feldern in der Vierfeldertafel)

Beim Vergleich der Beurteiler 2 und 3 ergaben sich prozentuale Übereinstimmungen von 33% bis 93% und Kappawerte von 0,07 bis 0,86 (siehe Tabelle 3). Für zwei Merkmale konnten keine Kappawerte berechnet werden. Die höchste überzufällige Übereinstimmung ergab sich beim Item „Klebriger Belag“ (0,86), bei „Auflösender Belag“ ergab sich eine gute Übereinstimmung (0,63). Eine moderate Übereinstimmung (0,48) ergab sich beim Merkmal „Zahneindrücke“, bei den übrigen Merkmalen lagen die Kappawerte unter 0,4.

Tabelle 3

Ergebnisse der ersten Pilotstudie (n = 15): Beurteiler 2 vs. Beurteiler 3

Item		Beurteiler 2 versus 3			Kappa	Signifikanz p-Wert
		Nein	Ja	%		
Zungenfarbe rot	nein	7	1	67%	0,31	0,18
	ja	4	3			
Zunge Blass	nein	7	6	60%	0,24	0,15
	ja	0	2			
Zungenfarbe bläulich	nein	5	0	40%	0,07	0,46
	ja	9	1			
Zungenfarbe rötlich- purpur	nein	14	0	93%	n.b.	n.b.
	ja	1	0			
Flecken	nein	4	0	33%	0,05	0,53
	ja	10	1			
Gedunsen	nein	5	6	60%	0,31	0,10
	ja	0	4			
Zahneindrücke	nein	7	4	73%	0,48	0,03
	ja	0	4			
Risse	nein	11	3	80%	0,33	0,09
	ja	0	1			
Dicker Belag	nein	9	6	60%	n.b.	n.b.
	ja	0	0			
Klebriger Belag	nein	6	0	93%	0,86	0,001
	ja	1	8			
Auflösender Belag	nein	13	0	93%	0,63	0,008
	ja	1	1			
Dünner Belag	nein	1	3	73%	0,19	0,42
	ja	1	10			

Übereinstimmungen fettgedruckt; n.b. = nicht berechenbar (bei zwei oder mehr leeren Feldern in der Vierfeldertafel)

Die Belagfarbe wurde von allen drei Beurteilern für alle 15 Zungenbilder als „weiß“ eingestuft. Die Ergebnisse sind nicht in die Tabellen aufgenommen.

4.2. Pilotstudie 2

Die 15 Patienten, deren Zungenfotos für die Pilotstudie 2 herangezogen wurde, waren im Schnitt 48 Jahre alt (Standardabweichung 7 Jahre) und zu 60% weiblich (siehe Tabelle 4). Die Mehrheit der Patienten (60%) hatte die Klinik zur Behandlung chronischer Schmerzsyndrome aufgesucht.

Tabelle 4

Alter, Geschlecht und Hauptdiagnosen der 15 Patienten

Alter	
- Mittelwert (Standardabweichung)	48 (7)
- Median (Minimum, Maximum)	47 (36, 61)
Geschlecht	
- weiblich	9 (60%)
- männlich	6 (40%)
Hauptdiagnosen	
- chronische Schmerzsyndrome	9 (60%)
- sonstige (jeweils n = 1)	6 (40%)

Beim Vergleich der Übereinstimmung ergaben sich wiederum in hohem Maße variable Kappa-Werte (-0,10 bis 1,00), allerdings lagen diese im Vergleich zur Pilotstudie 1 meist deutlich höher (Tabelle 5). Die Merkmale „Zungenspitze blass“ und „Zungenspitze bläulich“ traten aus Sicht beider Beurteiler nicht auf; daher konnten ebenso wie für das Merkmal „Zungenränder bläulich“ (das nur aus Sicht von Beurteiler 2 einmal auftrat) keine Kappa-Werte berechnet werden. Kappa-Werte von 1,00 ergaben sich für die Merkmale „Zungenkörper rot,“ „Zunge gedunsen,“ „Zahneindrücke,“ „Risse,“ „klebriger Belag,“ „sich auflösender Belag“ und für die Belagfarbe.“ Weniger hohe überzufällige Übereinstimmungen zeigten sich bei der Belagdicke, bei „Flecken“ und einzelnen Farbbeurteilungen („Zungenränder rot,“ „Zungenränder blass,“ „Zungenkörper blass“ und „Zungenkörper bläulich“).

Tabelle 5

Ergebnisse der zweiten Pilotstudie (n = 15)

Item		Beurteiler 1 vs. 2		%	Kappa	Signifikanz p-Wert
		Nein	Ja			
Zungenspitze rot	nein	1	1	93%	0,63	0,008
	ja	0	13			
Zungenspitze blass	nein	15	0	100%	n.a.	n.a.
	ja	0	0			
Zungenspitze bläulich	nein	15	0	100%	n.a.	n.a.
	ja	0	0			
Zungenränder rot	nein	12	2	87%	0,44	0,04
	ja	0	1			
Zungenränder blass	nein	5	8	47%	0,14	0,24
	ja	0	2			
Zungenränder bläulich	nein	14	1	93%	n.b.	n.a.
	ja	0	0			
Zungenkörper rot	nein	14	0	100%	1,00	<0,001
	ja	0	1			
Zungenkörper blass	nein	6	6	60%	0,29	0,11
	ja	0	3			
Zungenkörper bläulich	nein	6	0	67%	0,39	0,06
	ja	5	4			
Flecken	nein	12	0	87%	0,44	0,04
	ja	2	1			
Zunge gedunsen	nein	9	0	100%	1,00	<0,001
	ja	0	6			
Zahneindrücke	nein	9	0	100%	1,00	<0,001
	ja	0	6			
Risse	nein	14	0	100%	1,00	<0,001
	ja	0	1			
Dicker Belag	nein	12	2	80%	-0,10	0,68
	ja	1	0			
Klebriger Belag	nein	8	0	100%	1,00	<0,01
	ja	0	7			
Auflösender Belag	nein	14	0	100%	1,00	<0,001
	ja	0	1			
Dünner Belag	nein	5	0	80%	0,61	0,01
	ja	3	7			
Belagfarbe		Gelb	Weiß	100%	1,00	<0,001
	gelb	2	0			
	weiß	0	13			

Übereinstimmungen fettgedruckt; n.b. = nicht berechenbar (bei zwei oder mehr leeren Feldern in der Vierfeldertafel)

4.3. Hauptstudie

Die Patienten, deren Zungenfotos zur Auswertung herangezogen wurden, waren zu 63% weiblich und im Mittel 53 Jahre alt (Standardabweichung 16 Jahre; siehe Tabelle 6). Mehr als die Hälfte (52%) der Patienten wurde in der Klinik wegen chronischer Schmerzsyndrome (z.B. chronische unspezifische Schmerzen der Lendenwirbelsäule) behandelt, 12% wegen anderer Erkrankungen des 9% Nervensystems, 9% wegen Magen-Darm-Erkrankungen und 27% wegen verschiedener sonstiger, meist chronischer Erkrankungen.

Tabelle 6

Alter, Geschlecht und Hauptdiagnosen der 101 Patienten

Alter	
- Mittelwert (Standardabweichung)	53 (16)
- Median (Minimum, Maximum)	52 (20, 87)
Geschlecht	
- weiblich	64 (63%)
- männlich	37 (37%)
Hauptdiagnosen	
- chronische Schmerzsyndrome	53 (52%)
- Erkrankungen des Nervensystems	12 (12%)
- Erkrankungen des Magen-Darmsystems	9 (9%)
- Sonstige	27 (27%)

Die Ergebnisse der Zungenbewertung sind in den Tabelle 7 bis 9 dargestellt. Beim Vergleich der Beurteiler 1 und 2 lagen die Kappa-Werte zwischen 0,22 und 0,89, die prozentuale Übereinstimmung zwischen 73% und 98% (siehe Tabelle 7). Bei den Items „Zahneindrücke,“ „Risse,“ „sich auflösender Belag“ und bei der Belagfarbe ergaben sich sehr gute Übereinstimmungen über den Zufall hinaus (Kappa 0,78 bis 0,89); die prozentuale Übereinstimmung lag zwischen 94% und 98%. Bei den Items „Zungenränder rot“ und „Zunge gedunsen“ wurde eine gute Übereinstimmung beobachtet (Kappa zwischen 0,62 und 0,63; prozentuale Übereinstimmung 83% und 88%). Bei den Items „Zungenkörper blass,“ „Zungenkörper bläulich,“ „klebriger Belag“ und „dünner Belag“ war die Übereinstimmung moderat (Kappa zwischen 0,41 bis 0,54; prozentuale Übereinstimmung zwischen 73% und 79%). Bei den übrigen Merkmalen lag der Kappa-Index unter 0,40.

Tabelle 7

Ergebnisse Hauptstudie (n = 101): Beurteiler 1 vs. Beurteiler 2

Item	Beurteiler 1 vs. 2			%	Kappa	Signifikanz p-Wert
	Nein	Ja				
Zungenspitze rot	nein	7	10	84%	0,37	<0,001
	ja	6	78			
Zungenränder rot	nein	67	13	88%	0,62	<0,001
	ja	2	19			
Zungenkörper rot	nein	82	15	84%	0,22	0,002
	ja	1	3			
Zungenkörper blass	nein	51	18	74%	0,43	<0,001
	ja	9	23			
Zungenkörper bläulich	nein	36	21	73%	0,46	<0,001
	ja	7	37			
Flecken	nein	63	2	75%	0,35	<0,001
	ja	24	12			
Zunge gedunsen	nein	31	2	83%	0,63	<0,001
	ja	16	52			
Zahneindrücke	nein	62	4	95%	0,89	<0,001
	ja	1	34			
Risse	nein	82	2	94%	0,78	<0,001
	ja	4	13			
Dicker Belag	nein	79	17	83%	0,31	<0,001
	ja	0	5			
Klebriger Belag	nein	49	13	73%	0,41	<0,001
	ja	15	24			
Auflösender Belag	nein	92	1	97%	0,78	<0,001
	ja	2	6			
Dünner Belag	nein	27	9	79%	0,54	<0,001
	ja	13	52			
Belagfarbe		Gelb	Weiß	98%	0,88	<0,001
	gelb	8	1			
	weiß	1	90			

Übereinstimmungen fettgedruckt

Für die Beurteiler 1 und 4 variierten die Kappa-Werte zwischen 0,14 und 0,81, die prozentuale Übereinstimmung zwischen 63% und 97% (Tabelle 8). Beim Item „Belagfarbe“ ergab sich eine sehr gute Übereinstimmung über den Zufall hinaus (Kappa 0,81) bei einer prozentualen Übereinstimmung von 97%. Die Übereinstimmung für das Items „Zahneindrücke“ war gut (Kappa 0,71; prozentuale Übereinstimmung 87%). Bei den Items „Zungenspitze rot,“ „Zungenränder rot,“ „Zungenkörper blass,“ „Zunge gedunsen,“ „Risse“ und „sich auflösender Belag“ wurde eine moderate Übereinstimmung mit Kappa-Werten von

0,43 bis 0,57 bei prozentualen Übereinstimmung zwischen 79% und 93% beobachtet. Bei den übrigen Merkmalen lag der Kappa-Index unter 0,40.

Tabelle 8

Ergebnisse Hauptstudie (n =101): Beurteiler 1 vs. Beurteiler 4

Item		Beurteiler 1 vs. 4		%	Kappa	Signifikanz p-Wert
		Nein	Ja			
Zungenspitze rot	nein	13	4	79%	0,43	<0,001
	ja	17	67			
Zungenränder rot	nein	73	7	84%	0,50	<0,001
	ja	9	12			
Zungenkörper rot	nein	92	5	93%	0,33	0,001
	ja	2	2			
Zungenkörper blass	nein	63	6	83%	0,57	<0,001
	ja	12	20			
Zungenkörper bläulich	nein	18	39	65%	0,14	0,07
	ja	7	37			
Flecken	nein	48	17	64%	0,19	0,06
	ja	20	16			
Zunge gedunsen	nein	25	8	80%	0,56	<0,001
	ja	12	56			
Zahneindrücke	nein	54	12	87%	0,71	<0,001
	ja	2	33			
Risse	nein	69	15	83%	0,54	<0,001
	ja	2	15			
Dicker Belag	nein	90	6	91%	0,26	0,006
	ja	3	2			
Klebriger Belag	nein	47	15	63%	0,17	0,07
	ja	23	16			
Auflösender Belag	nein	89	4	93%	0,55	<0,001
	ja	3	5			
Dünner Belag	nein	21	15	71%	0,35	<0,001
	ja	15	50			
Belagfarbe		Gelb	Weiß	97%	0,81	<0,001
	gelb	7	2			
	weiß	1	90			

Übereinstimmungen fettgedruckt

Für die Beurteiler 2 und 4 ergaben sich Kappa-Werte zwischen 0,18 und 0,83 bei prozentualen Übereinstimmungen zwischen 71% und 98% (siehe Tabelle 9). Beim Item „Belagfarbe“ ergab sich eine sehr gute Übereinstimmung über den Zufall hinaus (Kappa 0,83; prozentuale Übereinstimmung 98%), beim Item „Zahneindrücke“ war die

Übereinstimmung gut (Kappa 0,69; prozentuale Übereinstimmung 86%). Bei den Items „Zungenspitze rot,“ „Zungenränder rot,“ „Zungenkörper blass,“ „Zungenkörper rot,“ „Zunge gedunsen,“ „Risse,“ „dicker Belag,“ „klebriger Belag,“ „sich auflösender Belag“ und „dünner Belag“ zeigte sich eine moderate Übereinstimmung (Kappa von 0,40 bis 0,59; prozentuale Übereinstimmung zwischen 71% und 94%). Bei den übrigen Merkmalen lag der Kappa-Index unter 0,40.

Tabelle 9

Ergebnisse Hauptstudie (n = 101): Beurteiler 2 vs. Beurteiler 4

Item		Beurteiler 2 vs. 4		%	Kappa	P
		Nein	Ja			
Zungenspitze rot	nein	12	1	81%	0,46	<0,001
	ja	18	70			
Zungenränder rot	nein	67	2	83%	0,56	<0,001
	ja	15	17			
Zungenkörper rot	nein	82	1	87%	0,42	<0,001
	ja	12	6			
Zungenkörper blass	nein	54	6	74%	0,51	<0,001
	ja	21	20			
Zungenkörper bläulich	nein	19	24	71%	0,36	<0,001
	ja	6	52			
Flecken	nein	62	25	70%	0,18	0,03
	ja	6	8			
Zunge gedunsen	nein	31	16	79%	0,56	<0,001
	ja	6	48			
Zahneindrücke	nein	52	11	86%	0,69	<0,001
	ja	4	34			
Risse	nein	68	18	79%	0,42	<0,001
	ja	3	12			
Dicker Belag	nein	78	1	84%	0,40	<0,001
	ja	15	7			
Klebriger Belag	nein	56	8	78%	0,51	<0,001
	ja	14	23			
Auflösender Belag	nein	90	4	94%	0,59	<0,001
	ja	2	5			
Dünner Belag	nein	24	16	73%	0,41	<0,001
	ja	12	49			
Belagfarbe		Gelb	Weiß	98%	0,83	<0,001
	gelb	8	1			
	weiß	0	90			

Übereinstimmungen fettgedruckt

5. Diskussion

5.1. Kurzzusammenfassung der Hauptergebnisse

Die vorliegende Untersuchungsreihe zeigt, dass eine Untersuchung der Reliabilität der morphologischen Beschreibung von Zungenmerkmalen möglich ist, aber mit erheblichen methodischen Schwierigkeiten verbunden ist. In den Voruntersuchungen zeigte sich, dass ohne eine angemessene Standardisierung der Beobachtungsbedingungen und Vorab-sprachen zwischen den Beurteilern über die Bewertung unklarer Befunde eine sinnvolle Studiendurchführung nicht möglich ist. Die Ergebnisse der Hauptstudie zeigen erwartungs-gemäß mit wenigen Ausnahmen für alle Items statistisch signifikante Übereinstimmungen über den Zufall hinaus. Allerdings ist das Ausmaß dieser Übereinstimmung (quantifiziert durch den Kappa-Index) bei der verwendeten Untersuchungsmethodik für die einzelnen Items sehr variabel und in vielen Fällen nur moderat. Bei Merkmalen wie zum Beispiel „Zahneindrücken“ ergaben sich hohe Übereinstimmungen während sich zum Beispiel die Farbbeurteilung der Zungen nicht zuletzt aus technischen Gründen bei der Bewertung am Computerbildschirm als schwierig erwies.

5.2. Diskussion der verwendeten Methoden

Das methodische Vorgehen in den Pilotstudien und der Hauptstudie wurde in enger Absprache mit den in der TCM und speziell der Zungendiagnostik ausgebildeten und in hohem Maße erfahrenen Ärzten entwickelt. Sowohl die beiden Pilotstudien wie auch die Hauptstudie wurden unter strikten Blindbedingungen durchgeführt, d.h. die Beurteiler bewerteten die Zungenfotografien streng unabhängig voneinander. Die Beurteiler hatten keinerlei weitere Informationen zu den Patienten. Da die Bilder aus einem großen Pool von mehreren tausend Fotografien ausgewählt wurden, erscheint es – vielleicht mit Ausnahme weniger sehr auffälliger Zungen - sehr unwahrscheinlich, dass sich die Beurteiler an die Patienten erinnerten. Dateneingabe und Auswertung wurden in einem unanhängigen Universitätsinstitut durchgeführt. Primäres Maß für die statistische Auswertung war nicht die (aufgrund des seltenen Auftretens vieler Merkmale) problematisch zu interpretierende prozentuale Übereinstimmung sondern der Kappa-Index. Während in den Pilotstudien aufgrund der begrenzten Ressourcen nur 15 Fotografien verwendet wurden, ist die Hauptstudie mit 101 Fotografien angemessen umfangreich und erlaubt somit eine

zuverlässige Aussage zur Übereinstimmung zwischen den teilnehmenden Beurteilern unter den gegebenen Studienbedingungen.

Trotz des systematischen Ansatzes trat bei der Untersuchungsreihe eine Vielzahl von Problemen auf, die zum Teil auch in zukünftigen Studien schwer zu lösen sein dürften. In den vorliegenden Studien erfolgte die Beurteilung auf der Basis von digitalen Zungenfotografien und nicht direkt am Patienten. Hierfür gab es mehrere Gründe: 1) sollte vermieden werden, dass die Beurteiler außer der Zunge selbst andere Informationen hatten bzw. die Patienten erkannten; 2) ist eine Durchführung derartiger Studien mit Fotografien logistisch in vieler Hinsicht (z.B. Rekrutierung, Ablauf, Verblindung) deutlich einfacher als direkt am Patienten; 3) erschien es anstrengenswert eine Methodik zu verwenden, die grundsätzlich auch an einem anderen Ort eine Wiederholung der Bewertung durch andere Beurteiler erlaubt. Die Verwendung von Fotografien und deren Beurteilung am Bildschirm erwies sich jedoch als problematisch. Obwohl die Zungenfotografien unter weitgehend standardisierten Bedingungen entstanden (gleiche Tageszeit, gleicher Raum, gleiche Sitzposition), war deren Qualität nicht einheitlich. So führen z.B. bereits geringe Lageveränderungen oder die Art und das Ausmaß wie der Patient die Zunge herausstreckt zu deutlichen Unterschieden in der Ausleuchtung. Ein grundsätzliches Problem bei Verwendung einzelner Zungenfotografien ist außerdem, dass im Gegensatz zur Beurteilung am realen Patienten, nur ein Eindruck aus einer einzigen Perspektive möglich ist. Dies führte z.B. dazu, dass die Beurteiler versuchten, den möglichen Einfluss von Schatten (also z.B. eine fälschliche Beurteilung der Zungenfarbe als dunkel rot) auf die Farbe abzuschätzen.

Die Entscheidung für eine Beurteilung am Bildschirm war getroffen worden, um eine mögliche weitere Verfälschung durch den Ausdruck der Fotografien zu vermeiden. Die Beurteilung am Bildschirm erwies sich jedoch durch den enormen Einfluss der Bildschirmqualität und des Blickwinkels als unbefriedigend. Um die Fehlerquellen zu minimieren, mussten daher alle Beurteilungen am gleichen Bildschirm und in Pilotstudie 2 und der Hauptstudie in einer extra angefertigten Vorrichtung zur weitgehenden Standardisierung des Blickwinkels erfolgen. Es kann aber nicht ausgeschlossen werden, dass verbleibende kleinere Unterschiede im Blickwinkel die Übereinstimmungsergebnisse ungünstig beeinflusst haben.

In der Praxis der Zungendiagnose spielt der Gesamteindruck der Zunge eine große Rolle. Eine detaillierte morphologische Beschreibung der Zunge erfolgt in der Routinedokumentation selten. Vielmehr erfolgt eine Interpretation der Zungenzeichen unter Einbezug der übrigen Informationen, Eindrücke und Befunde am Patienten. In der vorliegenden Studien-

serie wurde der Großteil dieses üblichen Vorgehens ausgeblendet. Um eine einfache Auswertung der Daten zu gewährleisten, waren auf dem Beurteilungsformular einzelne Merkmale als vorhanden oder nicht vorhanden zu beurteilen. Da die Auswahl der Zungenfotografien zufällig erfolgte, ist davon auszugehen, dass eine erhebliche Zahl uneindeutiger Befunde selektiert wurde. In der ersten Pilotstudie erfolgten keinerlei Absprachen, wie „Grenzwertbefunde“ zu bewerten seien. In der zweiten Pilotstudie und in eingeschränktem Ausmaß nochmals vor der Hauptstudie erfolgten mündliche Absprachen auf der Basis von Bildern aus den jeweiligen Vorstudien. Derartige „begrenzte“ Absprachen erscheinen für eine sinnvolle Studiendurchführung unumgänglich. Allerdings erfolgten die Festlegungen in den vorliegenden Studien nicht in schriftlicher Form.

Schließlich ist zu betonen, dass der gewählte Ansatz nur von begrenzter klinischer Relevanz ist und die gesamte Studienserie lediglich als Versuch eines ersten Ansatzes zu sehen ist, diagnostische Forschung zur Zungenbeurteilung durchzuführen.

5.3. Stand der Forschung zur Beschreibung der Zungenmorphologie im Rahmen der TCM

Wie bereits in der Einleitung angemerkt, scheint der Untersuchung der Reliabilität der Zungendiagnose bisher kaum Aufmerksamkeit geschenkt worden zu sein. Entsprechende Arbeiten aus China konnten trotz entsprechender Bemühungen über chinesische Fachleute nicht identifiziert werden. Die beiden nur als Abstracts publizierten westlichen Studien können nur als Pilotstudien interpretiert werden. Biltstein und Bonzak (2006) ließen drei erfahrene Akupunkteure 31 Merkmale bei 20 Zungenbildern beurteilen. Die ermittelten Kappawerte lagen zwischen 0,21 und 0,6. Unbefriedigende Kappawerte betrafen besonders Beurteilungen der Zungenspitze. Gareus et al. (2005) versuchten in einem ersten Schritt möglichst optimale und standardisierte Bedingungen für die Zungenfotografien zu erreichen, z.B. um einen möglichst immer identischen Lichteinfall zu gewährleisten. Zwei Beurteiler beurteilten dann den Zungenkörper und den Belag direkt am Patienten sowie am Computerbild und nach einer längeren Zeitspanne nochmals das Computerbild. Die Zahl der beurteilten Bilder und Details zur Art der Beurteilung sind nicht berichtet. Die Kappa-Indices beim Vergleich von Computerbild und Direktbeurteilung am Patienten waren 0,35 bzgl. Zungenkörper und 0,34 bzgl. Zungenbelag beim ersten und 0,65 bzw. 0,62 beim zweiten Beurteiler. Die Werte für die Test-Retest Reliabilität betrugen 0,53 bzw. 0,48 für den ersten und 0,65 bzw. 0,62 für den zweiten Beurteiler.

Über das Internet konnte ein Artikel gefunden werden, der eine Übersicht über den Stand der Zungendiagnoseforschung in China gibt (Zhang, ohne Angabe einer Jahreszahl, <http://www.tcminter.net/Artikel/Zungendiagnose.htm>). Demnach untergliedert sich die Forschung in diesem Bereich in China in folgende Bereiche: „Die Systematisierung der schriftlichen Informationen zur Zungendiagnose, die Untersuchung von gesunden Probanden zur Aufstellung eines Standards, die experimentelle Erforschung der Zungendiagnose, die Methodik der Diagnose des Zungengrundes, die Erforschung des Zungenfarbgrades, die Untersuchung der Zungenwärme mit Infrarotmessungen, die Entwicklung von Diagnosegeräten zur Zungendiagnose und schließlich die experimentelle und klinische Untersuchung der Zungenveränderung bei Krankheiten.“

Von besonderem Interesse für die vorliegende Arbeit sind in der Übersicht beschriebene Untersuchungen an mehreren tausend gesunden Probanden. Dabei wurde offensichtlich versucht, „Normalbefunde“ der Zunge zu ermitteln. Wie in der vorliegenden Dissertation scheinen dabei primär morphologische Merkmale erfasst worden zu sein. Leider zitiert der Übersichtsartikel keine Quellenangaben für die einzelnen Studien, so dass es nicht möglich war, die Methodik genauer zu analysieren. So bleibt unter anderem unklar, ob die Bewertung jeweils durch mehr als einen Beurteiler erfolgte. 89% der Probanden zeigten „eine feucht glänzende, klare und bewegliche Zunge.“ 57% der Zungen wurden als blassrot, 24% als rot klassifiziert. Mit höherem Alter nahm das Auftreten blassroter Zungen ab, während der Anteil violetter Zungen zunahm. Blassweiße Zungen fanden sich gehäuft bei jungen Frauen bis hin zum mittleren Alter, während bei jungen und Männern mittleren Alters häufiger ins rote tendierende Zungen oder solche mit roter Spitze beobachtet wurden, was in Relation zu den Rauchgewohnheiten stand. 54% der Zungenbeläge wurden als „überwiegend dünn-weiß“, 18% das „dünn-klebrig-weiß“ klassifiziert. Dünn-gelbe Beläge fanden sich bei 8% und dünn-klebrig-gelbe bei 14%. Andere Belagfarben wurden bei insgesamt 8% registriert. Mit zunehmendem Lebensalter verringerte sich die Anzahl der dünnen Beläge und die Anzahl der klebrigen Beläge nahm zu.

Von Interesse für die vorliegende Dissertation sind auch die von Yuan (2002) berichteten Vergleichsuntersuchungen an deutschen und chinesischen Patienten. So scheinen „deutsche Zungen“ im Allgemeinen etwas dunkler und röter zu sein als „chinesische Zungen“. Zungenfurchen, Zungenrisse und Zahneindrücke wurden bei deutschen Patienten häufiger beobachtet als bei chinesischen Patienten. Andererseits waren die Zungenbeläge deutscher Patienten in der Regel geringer ausgeprägt. Auch geschlechtsspezifisch scheinen

besondere Unterschiede zu bestehen. So war bei deutschen Frauen die Zungenfarbe an den Zungenrändern und der Zungenspitze im Schnitt dunkler als bei chinesischen Frauen.

Arbeiten zur Untersuchung der Reliabilität zitieren weder Zhang (<http://www.tcminter.net/Artikel/Zungendiagnose.htm>) noch Yuan (2002). Auch wenn nicht ausgeschlossen ist, dass derartige Arbeiten vereinzelt existieren, ist offensichtlich, dass in China andere Aspekte im Vordergrund des Forschungsinteresses stehen. Trotz der erheblichen Verbreitung der Zungendiagnose scheint es auch in anderen Ländern keine entsprechenden Aktivitäten zu geben.

5.4. Interpretation der Ergebnisse

Unabhängig von der Frage, ob die Zungendiagnose eine wissenschaftlich sinnvolle Methode darstellt und letztlich dazu beiträgt, dass individuelle Patienten möglichst erfolgreich behandelt werden können, konnte aufgrund des empirischen und vergleichsweise systematischen Charakters der Zungenbeurteilung davon ausgegangen werden, dass bei der Beschreibung morphologischer Merkmale überzufällig Übereinstimmungen auftreten. Dies konnte in der vorliegenden Studienserie eindeutig bestätigt werden. Die Ergebnisse sind mit denen der zwei anderen, als Abstracts publizierten Untersuchungen gut vereinbar (Blitstein und Bonzak, 2006; Gareus et al, 2005).

Die meist geringen überzufälligen Übereinstimmungen in der ersten Pilotstudie machen deutlich, dass a) methodische Probleme w.z.B. die Farbbeeinflussung durch den Blickwinkel auf den Computerbildschirm die Ergebnisse fundamental beeinflussen und b) eine fehlende Absprache über die Klassifizierung von grenzwertigen Befunden zu einer ungerechtfertigt negativen Beurteilung der Reliabilität der morphologischen Beurteilung führen kann. Die Behebung dieser Probleme führte in der zweiten Pilotstudie und in der Hauptstudie zu deutlich besseren, plausibel erscheinenden Übereinstimmungen. Höhere Übereinstimmungen mit Kappawerten durchgehend über 0,6 sind vermutlich nur durch ein intensives Beurteilertraining zu erreichen. Es stellt sich aber die Frage, in wie weit dies jedoch dann noch die Anwendung der Zungendiagnose in der Praxis wieder spiegelt.

Es ist schwer abschätzbar, inwieweit die meist nur moderaten überzufälligen Übereinstimmungen dazu führen, dass die *Interpretation* der Zungenzeichen sich unterscheidet. Hier erscheinen sowohl höhere Übereinstimmungen (durch die synthetische Beurteilung der

vielen „kleinen“ Zeichen) wie auch geringere (durch eine „Aufaddierung“ kleiner Fehler oder die zunehmende Komplexität) möglich.

5.5. Schlussfolgerung und Empfehlungen für zukünftige Forschungsvorhaben

Die vorliegenden Ergebnisse deuten darauf hin, dass die Beurteilung morphologischer Merkmale im Rahmen der Zungendiagnose eine befriedigende Reliabilität haben dürfte. Eine weitergehende Schlussfolgerung ist angesichts der geringen Zahl insgesamt vorliegender Untersuchungen, aufgrund der methodischen Schwierigkeiten und aufgrund der unklaren Übertragbarkeit der Ergebnisse auf andere Personen, die TCM praktizieren, nicht vertretbar.

Angesichts der Verbreitung der TCM und der Relevanz der Zungendiagnostik im Rahmen dieses Therapiesystems erscheint eine weitere Erforschung wünschenswert. Studien zur Reliabilität der morphologischen Beurteilung sind hier sinnvoll, wenn auch nur als erster Schritt hin zu Untersuchungen, die die Reliabilität der Interpretation der Zungenzeichen überprüfen sowie Studien, die sich mit der Validität der Zungendiagnose im Rahmen der TCM auseinander setzen.

Als Fortsetzung der vorliegenden Studienserie erscheint eine ganze Reihe von Untersuchungen möglich. Derzeit wird in der TCM-Klinik in Zusammenarbeit mit dem Zentrum für naturheilkundliche Forschung der Technischen Universität München eine Folgestudie durchgeführt. Wiederum soll die Reliabilität der morphologischen Beschreibung untersucht werden. Allerdings erfolgt die Beurteilung diesmal mit ausgedruckten Photographien. Dies soll die Probleme bei der Beurteilung am Computerbildschirm vermeiden und außerdem ermöglichen, dass auch außerhalb der TCM-Klinik weitere Beurteiler einbezogen werden können. Ein weiterer wichtiger Schritt zur Standardisierung und Übertragbarkeit der Vorgehensweise ist die Erstellung eines Manuals, in dem für jedes einzelne Merkmal Beispielfotos gezeigt werden. Insbesondere werden Empfehlungen gegeben, wie „Grenzwertbefunde“ beurteilt werden sollen. Als erster Schritt für weitergehende Untersuchungen wird auch die Interpretation der Zungenzeichen als Freitext erfasst.

Sollte sich die verwendete Methodik bewähren, könnte auch die Intra-Rater-Reliabilität untersucht werden, d.h. es könnte geprüft werden, ob ein Beurteiler nach einer längeren Zeit eine Zunge gleich beurteilt. Für die Frage, ob Studien zur Interpretation der Zungenzeichen überhaupt mit Fotografien durchgeführt werden können, ist ein Vergleich der Beurteilungen

direkt am Patienten notwendig. Die oben genannten Vorarbeiten müssen hierfür jedoch erfolgreich absolviert sein.

Angesichts der offensichtlich umfangreichen Bemühungen in China zur Digitalisierung und Objektivierung der Dokumentation (Chiu, 2000; Wei et al., 2002; Zhou et al., 2004) muss versucht werden, eine bessere Übersicht über den Stand der Forschung zu erlangen und eventuell vorhandene Erkenntnisse für weitere Untersuchungen auch im Westen zu nutzen.

6. Literaturverzeichnis

- Blitstein R, Bonzak S. Interobserver reliability in traditional Chinese medicine tongue observation. J Alternat Complement Med 2006;12:206.
- Blunck A, Focks C, Höll A, Hertzner D, Kubny M. Theoretische Grundlagen. In Focks C, Hillenbrand N (Hrsg). Leitfaden Chinesische Medizin. München: Urban & Fischer, 5. Auflage, 2006, 45-146.
- Chiu CC. A novel approach based on computerized image analysis for traditional Chinese medical diagnosis of the tongue. Comput Methods Programms Biomed 2000;61:77-89.
- Crivelli L, Ferrari D, Limoni C. Inanspruchnahme von 5 Therapien der Komplementärmedizin in der Schweiz. Statistische Auswertung auf der Basis der Daten der Schweizer Gesundheitsbefragung 1997 und 2002. Verfügbar im Internet unter: <http://www.bag.admin.ch/themen/krankenversicherung/00305/02363/index.html?lang=de#> (Zugang zuletzt geprüft am 27.3.2007).
- Eisenberg DM, Davis RB, Ettner SL, Appel S, Wilkey S, Van Rompey M. Trends in alternative medicine use in the United States, 1990 -1997. Results of a follow-up national survey. JAMA 1998;280:1569-75.
- Fleiss JL. Statistical methods for rates and proportions (2nd ed). New York: Wiley, 1981.
- Fletcher RW, Fletcher SW. Clinical epidemiology - the essentials. Philadelphia: Lippincott, Williams & Wilkins, 2005, S 36.
- Gareus IK, Tan L, Lüdtke R, Niggemeier C, Li FM, Bäcker M, Lehmann C, Klose P, Spahn G, Dobos GJ. Introducing a computer-assisted, digital tongue-imaging device to standardise and evaluate traditional Chinese tongue diagnosis. Focus on Alternative and Complementary Therapies 2005;10 (Suppl 1):20-21.
- Härtel U, Volger E. Inanspruchnahme und Akzeptanz klassischer Naturheilverfahren und alternativer Heilmethoden in Deutschland - Ergebnisse einer repräsentativen Bevölkerungsstudie. Forsch Komplementärmed Klass Naturheilkd 2004;11:327-34.
- Jensen K, Abel U. Methodik diagnostischer Validierungsstudien. Fehler in der Studienplanung und Auswertung. Med Klin 1999;94:522-8.
- Kirschbaum B. Atlas und Lehrbuch der Chinesischen Zungendiagnostik (in 2 Bänden). Kötzing: Verlag für Ganzheitliche Medizin Dr. Erich Wühr Verlag, 2002.

- Lao L. Traditional Chinese medicine. In: Jonas WB, Levin JS (Hrsg). Essentials of complementary and alternative medicine. Philadelphia: Lippincott Williams & Wilkins, 1999, 217-232.
- Lu AP, Jia HW, Xiao C, Lu QP. Theory of traditional Chinese medicine and therapeutic method of diseases. World J Gastroenterol 2004;10:1854-1856.
- Maciocia G. Zungendiagnose in der Chinesischen Medizin. Uelzen: Medizinisch Literarische Verlagsgesellschaft, 2000.
- Maret A, Tang J, Roth L, Wiesemann G, Tian L, Focks C. Diagnostik. In Frocks C, Hillenbrand N (Hrsg). Leitfaden Chinesische Medizin. München: Urban & Fischer, 5. Auflage, 2006, 166-253.
- Melchart D, Wichmann B, Polonius D, Saller R Implementierung und Erprobung einer multifunktionalen elektronischen Patientenakte für Routine und Evaluation der Patientenversorgung. Praxis 2001;90: 1587-1593.
- Neeb G. TCM-Studium in China und Taiwan. In Frocks C, Hillenbrand N (Hrsg). Leitfaden Chinesische Medizin. München: Urban & Fischer, 5. Auflage, 2006, 5-19.
- Porkert M. Die Chinesische Medizin. 2. Auflage. Düsseldorf: Econ, 1986.
- Renckens CNM. Alternative treatments in reproductive medicine: much ado about nothing. Hum Reproduct 2002;17:528-533.
- Tindle HA, Davis RB, Phillips RS, Eisenberg DM. Trends in use of complementary and alternative medicine by US adults: 1997-2002. Altern Ther Health Med 2005;11:42-9.
- Unschuld PU. Chinesische Medizin. 2. Auflage München: CH Beck, 2003.
- Wei BC, Shen LS, Wang YQ, Wang YG, Wang AM, Zhao ZX. A digital tongue image analysis instrument for traditional Chinese medicine. Zhongguo Yi Liao Qi Xie Za Zhi 2002;26:164-9. (Artikel in chinesischer Sprache)
- WHO Global Atlas of Traditional, Complementary and Alternative Medicine. Kobe: World Health Organization; Centre for Health Development, 2005.
- Yuan H. Traditionelle Chinesische Zungendiagnostik. Wiesbaden: Ullstein Mosby, 1996.
- Zhang BL. Erforschung der Systematisierung und Objektivierung der TCM-Zungendiagnose. Verfügbar im Internet unter <http://www.tcminter.net/Artikel/Zungendiagnose.htm>
- Zhou Y, Yang J, Shen L. Methodological study on digitalization of tongue image in traditional Chinese medical diagnosis. Sheng Wu Yi Xue Gong Cheng Xue Za Zhi 2004;21:917-20. (Artikel in chinesischer Sprache)

7. Zusammenfassung

Hintergrund und Fragestellung: Die Traditionelle Chinesische Medizin (TCM) ist eine auch außerhalb Asiens weit verbreitete komplementärmedizinische Methode. Die Beurteilung der Zunge ist ein wesentlicher Bestandteil der Diagnostik und ist unverzichtbar für eine korrekte Syndromdiagnose nach TCM. Forschung zur Zungendiagnostik wurde außerhalb Chinas bisher nur in sehr begrenztem Rahmen durchgeführt. In der dieser vorliegenden Dissertation zugrunde liegenden Studienserie sollten erste Schritte für eine wissenschaftlich angemessene Untersuchung der Reliabilität der Zungendiagnostik im Rahmen der TCM unternommen werden. Untersucht wurde die Frage, inwieweit unterschiedliche Beurteiler bei der morphologischen Beschreibung für die Zungendiagnostik wichtiger Merkmale übereinstimmen.

Methoden: In einer Klinik für Traditionelle Chinesische Medizin in Kötzing (Deutschland) wurden zwei Pilotstudien und eine größere Untersuchung (Hauptstudie) durchgeführt. Beide Pilotstudien und die Hauptstudie waren als standardisierte, deskriptive Beurteilungen von Zungenmerkmalen durch drei (erste Pilotstudie und Hauptstudie) bzw. zwei (zweite Pilotstudie) unabhängige, verblindete Beurteiler angelegt. Aus einer Datenbank mit mehreren tausend, unter weitgehend standardisierten Bedingungen ausgenommenen Digitalfotografien von Zungenbefunden von Klinikpatienten wurden für die beiden Pilotstudien jeweils 15 Bilder und für die Hauptstudie 101 Bilder zufällig gezogen. Die Beurteilung erfolgte am Computerbildschirm mit Hilfe von Formularen, auf denen die Beschreibung der morphologischen Merkmale (je nach Studie 13 bis 18 Items zu Farbe, Flecken, Form und Belag) in standardisierter Form festgehalten wurde. Wichtigstes Zielkriterium war die Übereinstimmung über die zufällig erwartete Übereinstimmung hinaus quantifiziert mit dem Kappa-Index. Kappawerte $< 0,40$ wurden als geringe überzufällige Übereinstimmung klassifiziert, Werte zwischen 0,40 und 0,59 als moderate Übereinstimmung, Werte zwischen 0,60 und 0,74 als gute und Werte von mindestens 0,75 als sehr gute Übereinstimmung.

Ergebnisse: In der ersten, mit drei Beurteilern durchgeführten Pilotstudie wurden in hohem Maße variable Kappawerte beobachtet (-0,15 bis 0,76), für zahlreiche Merkmale zeigten sich jedoch nur geringe bis mäßige Übereinstimmungen zwischen den Beurteilerpaaren. Als Ursache hierfür wurden vor allem Änderungen der Farbdarstellung am Computerbildschirm durch unterschiedliche Blickwinkel und fehlende Abstimmung bei der Klassifizierung von grenzwertigen Befunden identifiziert. In der zweiten, mit zwei Beurteilern durchgeführten Pilotstudie, in der diese Mängel so weit möglich behoben wurden, zeigten sich zwar wiederum in hohem Maße variable Kappawerte (-0,10 bis 1,00), für 7 von 18 erhobenen

Merkmale war die überzufällige Übereinstimmung jedoch sehr gut. In der mit drei Beurteilern (Beurteiler 1, 2 und 4) durchgeführten Hauptstudie lagen die Kappa-Werte für das Beurteilerpaar 1 und 2 zwischen 0,22 und 0,89, für das Beurteilerpaar 1 und 4 zwischen 0,14 und 0,81 und das Beurteilerpaar 2 und 4 zwischen 0,18 und 0,83. In 39 von 42 Vergleichen ging die Übereinstimmung in statistisch signifikanter Weise über das zufällige zu erwartende Ausmaß hinaus, in 35 Vergleichen war der p-Wert $< 0,001$. Gute bis sehr gute Übereinstimmungen ($\text{Kappa} > 0,59$) zwischen allen Beurteilern wurden für die Merkmale Zahneindrücke und Belagfarbe gefunden. Durchgehend geringe bis mäßige überzufällige Übereinstimmungen zeigten sich bei der Beurteilung von Flecken, einzelnen Farbbeurteilungen und bei der Frage, ob der Belag dick sei.

Diskussion: Die vorliegende Untersuchungsreihe zeigt, dass eine Untersuchung der Reliabilität der morphologischen Beschreibung von Zungenmerkmalen möglich ist, aber mit erheblichen methodischen Schwierigkeiten verbunden ist. In den Voruntersuchungen zeigte sich, dass ohne eine angemessene Standardisierung der Beobachtungsbedingungen und Vorabsprachen zwischen den Beurteilern über die Bewertung unklarer Befunde eine sinnvolle Studiendurchführung nicht möglich ist. Die Ergebnisse der Hauptstudie zeigen erwartungsgemäß mit wenigen Ausnahmen für alle Items statistisch signifikante Übereinstimmungen über den Zufall hinaus. Allerdings ist das Ausmaß dieser Übereinstimmung (quantifiziert durch den Kappa-Index) bei der verwendeten Untersuchungsmethodik für die einzelnen Items sehr variabel und in vielen Fällen nur moderat. Die Ergebnisse deuten darauf hin, dass die Beurteilung morphologischer Merkmale im Rahmen der Zungendiagnose eine befriedigende Reliabilität haben dürfte. Eine weitergehende Schlussfolgerung ist angesichts der geringen Zahl insgesamt vorliegender Untersuchungen, aufgrund der methodischen Schwierigkeiten und aufgrund der unklaren Übertragbarkeit der Ergebnisse auf andere Personen, die TCM praktizieren, nicht vertretbar.

8. Summary

Background and objective: Traditional Chinese Medicine (TCM) has gained considerable popularity in Western countries in recent years. The assessment of tongue signs is a crucial diagnostic tool in TCM and considered mandatory for establishing an adequate syndrome diagnosis according to TCM principles. Tongue diagnosis has rarely been investigated in empirical studies outside China. In the series of studies reported in this MD thesis it was tried to develop adequate methods to investigate the reliability of tongue diagnosis. We investigated to which extent independent observers agreed beyond chance in the description of morphological characteristics of the tongue relevant to TCM.

Methods: Two pilot studies and a larger confirmatory study were performed in a hospital for traditional Chinese medicine located in Kötzing, Germany. All three studies included a standardized, descriptive assessment of morphological tongue characteristics by three (pilot study 1 and confirmatory study) or two (pilot study 2) independent, blinded raters. From a pool of several thousand digital tongue pictures taken under standardised conditions 15 pictures were randomly chosen for each of the pilot studies and 101 for the confirmatory study. The raters assessed the photographs displayed on a computer screen using standardized, pretested forms which included between 13 and 18 items on colour, spots, forms and coating of the tongue. Main outcome measure was assessment over chance quantified by the Kappa-Index. Kappa values below 0.4 were considered as low agreement beyond chance agreement, values between 0.4 and 0.59 as moderate agreement, values between 0.6 and 0.74 as good, and higher values as excellent agreement.

Results: In the first pilot study Kappa values varied to a great extent between the three raters (range -0.15 to 0.76). However, agreement beyond chance was low or moderate for most items. Major reasons for disagreements seemed changes of colours on the computer screen depending on the angle of the rater and the screen, and the lack of guidelines on how to rate borderline tongue signs. In the second pilot in which only two raters participated these drawbacks were corrected. While Kappa values still varied strongly (-0.1 to 1.00), agreement over chance was good to excellent for seven of 18 items. In the confirmatory study with three raters (rater 1, 2 and 4) Kappa values ranged from 0.22 to 0.89 between rater 1 and rater 2, from 0.14 to 0.81 between rater 1 and rater 4, and from 0.18 to 0.83 for rater 2 and rater 4. In 39 of 42 tests the observed agreement exceeded that expected by chance in a statistically significantly manner, in 35 cases p-values were below 0.001. Good to excellent agreement between all raters was observed for tooth imprints and the colour of the coating, while

agreement on spots, specific colour characteristics and thick coating was consistently low to moderate.

Discussion: The performed series of studies shows that it is possible to scientifically investigate the reliability of describing morphological tongue characteristics relevant to TCM. However, the studies also show that the methodology of such investigations is prone to a number of problems. The pilot studies demonstrate that it is not possible to obtain meaningful results without adequate standardisation of the conditions and without guidelines on how to classify the frequent borderline findings. The findings of the confirmatory study indicate that raters in general agree clearly beyond chance, however, with the methods applied kappa values varied considerably and were often not fully satisfactory. Our findings suggest that the morphological assessment of tongue signs in the framework of TCM has some reliability. More definitive conclusions are currently not possible due to the small amount of studies available, due to the considerable methodological problems which are faced in such studies and due to the unclear generalisability of the findings.

Danksagung

Herrn Prof. Dr. R. Saller und PD Dr. D. Melchart danke ich sehr herzlich für die Überlassung des Themas und sorgfältige Betreuung der Dissertation. Dem Leiter des Zentrums für naturheilkundliche Forschung (Znf), Herrn Dr. Melchart, gebührt mein aufrichtiger Dank für die mehrjährige gute und vertrauensvolle Zusammenarbeit.

Bei meinen Arbeitskollegen, den Herr PD Dr. K. Linde, möchte ich mich sehr herzlich für die Betreuung, die konstruktiven Diskussionen, die vielen Ratschläge bei der Fertigstellung der Arbeit bedanken.

Den Ärzten der TCM Klinik Kötzing, vor allem dem Chefarzt, Herrn Dr. Hager, und dem Leiter der chinesischen Ärztedelegation, Herrn Prof. Dai, bin ich zu besonderem Dank verpflichtet. Ohne deren großes Engagement hätte die Studie und damit letztlich auch die Dissertation nicht erfolgreich durchgeführt werden können.

Meinem geliebten Mann Peter danke ich für die unermüdliche Ermunterung und Unterstützung und nicht zuletzt für die Geduld, die er über Jahre hinweg für meine Doktorarbeit aufgebracht hat.

Schließlich danke ich meiner geliebten Mutter, was sie mir im Laufe meines Lebens Gutes getan hat, und besonders auch dafür, dass sie mir inneren Mut gegeben hat, die Schwierigkeiten bei der Erstellung meiner Doktorarbeit in einer fremden Sprache durchzustehen.